图书情报工作
第 65 卷 第 2 期 2021 年 1 月

面向中医诊疗知识库的医案元数据模型构建研究

■ 李璐[1] 马捷[1,2] 孙恒宇[3] 王珏[1]
[1] 吉林大学管理学院 长春 130022 [2] 吉林大学信息资源研究中心 长春 130022
[3] 吉林省中医药科学院第一临床医院 长春 130022

摘 要：[目的/意义] 在中医医案管理过程中存在中医医案元数据不统一、精度不足等问题，不利于中医医案知识库的信息检索、知识共享和挖掘，研究建立规范的中医医案元数据模型，对于支持中医医案的深度多维检索、知识共享和挖掘具有重要作用。[方法/过程] 通过文献调研法对现用中医相关标准规范、元数据标准进行总结；通过网络调查法收集中医医案实例，结合内容分析法对医案内容进行分析，初步界定医案元数据元素；再通过实地调查法，在中医诊疗现场对中医诊疗过程进行观察和访问，对初步界定的医案元数据元素进行优化，得到最终医案元数据元素集。[结果/结论] 构建了面向中医诊疗知识库的医案元数据模型，全面描述中医诊疗过程，为统一中医医案元数据提供参考。

关键词：元数据 元数据模型 中医医案 中医诊疗知识库 中医知识组织
分类号：G250
DOI：10.13266/j.issn.0252－3116.2021.02.001

中医学是我国特有的文化和科学遗产，中医医案作为中医学的传承工具，体现了医家临床实践和辨证论治的宝贵经验，是历代学者学习前代医者学术思想及临证经验的重要途径[1]。随着医案数据增长及我国中医学现代化研究的深入，医案信息管理逐渐由传统的纸质管理转变为新兴的信息技术及数据库管理。中医诊疗数据库和知识库是能够实现对中医医案进行存储、共享、挖掘和分析等功能的数据和知识平台，相对于诊疗数据库，诊疗知识库更重视知识组织、知识挖掘功能的开发。目前，国内已建设了一批中医医案信息数据库和知识库，如："中医药基础数据库系统"[2]、"历代医案分析统计系统"[3]、"中国中医药文献数据库"[4]等。但是这些数据库建成后，使用率不理想，互不联通，海量的医案资源没有得到有效的开发和利用，存在的主要问题有：①这些数据库中包含的医案大多是历代医案数据，所记载的医案具有其自身的个性化特点，医案元数据格式不统一；②现有医案数据库或知识库的元数据不够详细，不能全面结构化地揭示中医诊疗过程中的细节，不利于中医医案的高精度检索和知识发现。针对以上问题，本文以面向中医诊疗知识库的医案元数据规范为目标，在系统化研究基础上，提出一套中医医案元数据模型，规范医案信息记录格式，形成深度揭示医案内容的基础，促进中医医案知识资源能够得到更好地开发和利用。

1 中医医案元数据研究现状

1.1 中医医案的内涵及发展

医案，又称为诊籍、脉案、方案、病案等[5]，是医生诊治病人的病情及治疗疾病时辨证、立法、处方用药的连续记录[6]。总体来说，中医医案是对诊疗过程最原始和可靠的记录，是中医诊疗疾病过程与实践经验的总结，它保存了中医治疗疾病的海量信息，是中医诊疗知识库的诊疗知识单元。

医案经过两千余年的发展，由最开始仅记录患者的基本信息、病证、病因病机、治法、疗效、预后等重要内容[7]，发展至近现代参考西医病案的"病历式"特征，包括患者基本信息、就诊时间、主诉、现病史、既往史，中医四诊情况，辅助检查结果，诊断及治疗意见等。但医案与病历相似却不尽相同，二者都是诊疗记录，病历是按照现代医院诊疗模式全面记录患者健康状况和疾病过程，医案是重点记录中医临证事实和诊疗思维活动[8]。二者的区别见表 1[9]。病历是医案的基础，

作者简介：李璐（ORCID：0000－0003－0460－1789），硕士研究生；马捷（ORCID：0000－0002－1471－2143），教授，博士，博士生导师，通讯作者，E-mail：m-lj-1@163.com；孙恒宇（ORCID：0000－0003－0882－7507），主治医师；王珏（ORCID：0000－0001－5272－7904），硕士研究生。
收稿日期：2020－05－19 修回日期：2020－07－30 本文起止页码：4－16 本文责任编辑：易飞

是对患者健康状况及治疗方法的记录，医案比病历记录更精细，同时还包括医生对病症辩证的过程。辩证过程是医生行医经验的隐性知识显性化的体现，大多存在于医案的按语中。

表1　医案与病历区别

项别	医案	病历
内容	重点突出	全面收集
表述	暂无模式	规范化，表格化
评述	多有按语	无按语
作用	研究，指导	查阅，追询

1.2　已有医学相关标准规范

中西医领域均有病历标准规范，但是医案的标准规范比较鲜见，由于病历是医案的基础，本文面向中医诊疗知识库的医案元数据模型构建就以现有的病历标准为基础，加入中医诊疗过程细节元数据，深入全面地表达医案内容。

1.2.1　中医中药元数据标准

作为拥有科学数据的大国，我国在中医学研究领域，已积累大量的科学研究与数据[10]，目前中国医学科学院、中国中医研究院等机构已联合制定《医药卫生科学数据共享元数据标准》和《中医药学科学数据共享元数据标准》，为中医数据库的建设与中医相关元数据的结构设计提供参考；国际上，近年来我国在ISO/TC 215(健康信息学技术委员会)开展了一系列面向传统医学的标准化工作[11]，目前已经发布了3项中医药信息标准，其中ISO/TS 17938《中医药文献元数据》为中医药学的文献资源提供了一套通用的描述元素[12]。

1.2.2　中医病历及临床数据标准规范

在中医病历标准规范方面，随着中西医结合的推进，中医诊疗过程的记录逐渐转变为"病历式"文本保存，我国出台了一系列关于中医的病历标准与规范，如：《中医病历书写基本规范》《中医、中西医结合病历书写指南》等。这些规范中包含了医生们诊疗的大体流程和部分元素；目前，信息技术飞速发展，许多纸质版信息已经转变为电子信息存档，以便信息的管理与保存。我国颁布了关于电子病历的标准与规范，如：《电子病历基本数据集标准》(WS 445－2014)等。这些标准不仅包括纸质病历包含的元素，还设计各类适合信息保存的数据结构以供参考；信息技术快速提升的同时，"健康中国"已成为当前我国重要发展战略之一，健康全民化是国家一直以来关注的焦点。在2009年卫生部发布了关于健康档案、卫生信息的一些文件，如《卫生信息数据集元数据规范(WS/T 305－2009)》等；为了保证元素来源的全面性，本文也参照了中医诊疗过程的相关书籍和模板，例如《中医诊断学》《诊断学大辞典》、"传统中医师承病案模板"等。

国际上，CDISC(临床数据交换标准协会)建立了一系列临床研究数据标准，用于临床数据的电子获取、交换、递交和归档[13]。我国中医药临床研究病例报告表的设计基于CDISC标准中的CDASH(临床数据获取协调标准)，同时增添了中医临床数据特色，例如中医诊断方面等内容，以促进中医药临床研究数据交换与共享，支持临床研究与临床工作流整合[14]。

1.3　中医医案元数据文献调研

我国医学元数据研究的水平目前还处于初期发展阶段，以"主题词＝中医 AND 关键词＝元数据"为检索式，限定截止时间为2020年4月23日，在中国知网中进行文献检索，获得结果仅有17篇文献，其中多为对中医医案文献元数据的研究，仅有4篇是与中医医案元数据相关的文献：2篇[15－16]是对中医医案数据库的需求与研究现状分析，指出对于数据复杂的中医医案，需要建立一套完善的元数据方案对医案信息统一描述，才能建立相同数据格式和数据结构的中医医案数据库；2篇[17－18]是关于病案数据库元数据的方案和结构设计，将病案数据库的结构分为医案基本信息元数据模块、疾病基本信息元数据模块、诊断、治疗、方药、错误操作、禁忌、西医相关信息、医嘱9个部分，并对这9个部分进行细化(见图1)，但仅是初步尝试，对诊疗过程知识的结构较为概括，元数据元素的细化不够充分。

综上所述，我国中医医案元数据的研究还处于起步阶段，学术界进行了初步的探索，远远不能满足中医诊疗知识组织和中医传统文化传承的需要。本文将结合这些问题，以病历(或电子病历)规范及已有元数据标准或结构为基础，根据医案实例将已有元数据的粒度细化，使录入中医诊疗知识库中的医案信息可以更全面深入地揭示中医诊疗知识，为更进一步的中医知识检索和挖掘提供基础和支撑。

2　中医医案元数据模型构建原则与流程

2.1　元数据模型构建方法

大多数元数据模型都是参照已有元数据标准，根据具体业务建立的，反映某领域具体业务的模型。例如：艺术品数字化元数据模型是在借鉴已有标准体系和研究框架的基础上，结合研究人员与管理人员的元数据使用需求而构建，将元数据根据描述内容和管理

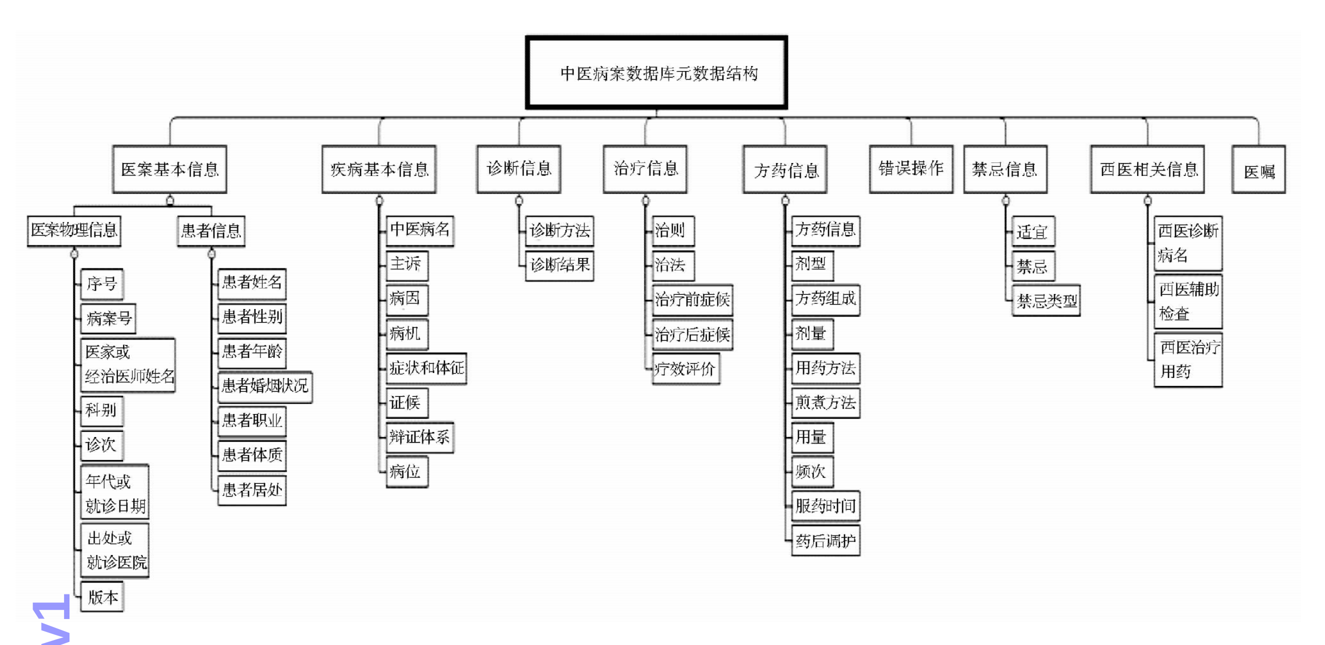


图 1　中医病案数据库元数据结构[16]

机制等方面功能不同进行了模块化划分[19]；数字博物馆元数据模型是对已有元数据标准进行改造，重新分类，建立动态的满足资源需求的元数据模型[20]；地质灾害应急信息元数据模型是根据实践中应急信息资源整理分类构建的[21]。可见，元数据模型构建方法一般为以下两种或这两种结合：一种是对已有标准规范的扩展或改造，一种是根据具体业务流程设计模型。本文提出的中医医案元数据模型将以上两种方法结合，根据已有标准规范进行元素扩展，再通过实际案例检验并优化元素，保证元数据模型准确实用，满足需求。

2.2　元数据模型扩展原则

元数据的扩展原则主要有两种[22]：一是横向扩展，即根据扩展原则增加一个元数据元素，但不能跟已有元素有任何语义上的重复；二是纵向扩展，即对当前元数据元素语义进行更加严格的限定，生成新的元数据修饰词[23]，强化描述的精深度。由于一层元素的修饰词无法精准描述该元素，纵向扩展可能增加元素修饰词的元素修饰词，在本文中，以一级元素、二级元素、三级元素表示。

2.3　中医医案元数据模型设计流程

本文中拟基于“元素初步界定——元素的验证与优化 ——元数据模型建立”三个步骤（见图 2）完成中医医案元数据元素的确定。首先通过对中医相关的标准规范和现有元数据标准进行调研，了解中医诊疗过程，确定核心元素及扩展的元素；再对网站中的医案进行筛选，运用内容分析法分析筛选出的医案，与上述标准调研确定的元素进行对比，查缺补漏，初步界定元素；其次为了保证元素的全面性与准确性，引入实地调查法，观察与记录中医医生的诊病过程，分析记录后对初步界定的元素进行优化；最后确定中医医案元数据元素集所有的元素。

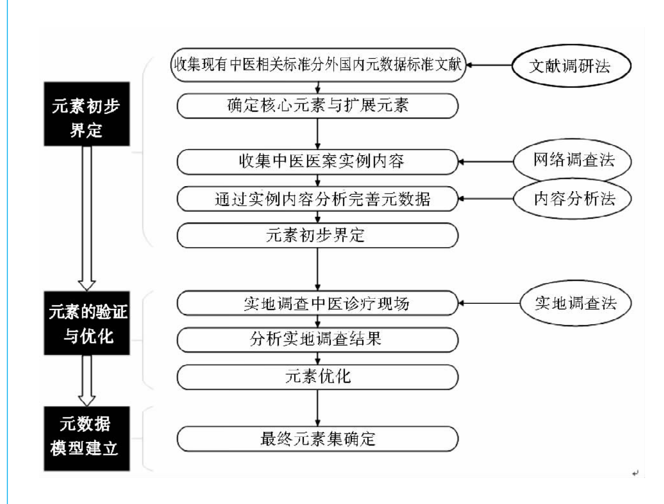


图 2　中医医案元数据模型设计流程

3　中医医案元数据元素初步界定

3.1　确定核心元素和扩展元素

参考上文中医医案（病历）相关标准规范，根据《中医病历书写规范》以及美国的 SOAP 药历书写格式，将医案元数据分为 6 个核心元数据：医案基本信息、患者基本信息、诊断过程、诊断评定、诊疗措施、分析总结。其中，诊断过程、诊断评定、诊疗措施三项是参照 SOAP 病历设定的。SOAP 病历是美国临床药师协会推荐的药历书写格式，S 代表 Subjective，即主观性

李璐，马捷，孙恒宇，等. 面向中医诊疗知识库的医案元数据模型构建研究[J]. 图书情报工作,2021,65(2):4 – 16.

资料,包括患者的主诉、病史、既往史等;O 代表 Objective,即客观性资料,包括患者临床各种医疗检验结果;由于 S 项与 O 项都是医生在诊疗过程中需要判断病情的依据,将二者合并为“诊断过程”元数据。A 代表 Assessment,即临床诊断以及对治疗过程的分析与评价,对应“诊断评定”元数据;P 代表 Plan,包括治法、用药指导及其他相关建议,即治疗方案,对应“诊疗措施”元数据。“分析总结”是根据“传统中医师承病案模板”最后的“按语”设定的,主要是对整个医案实例参照的药方或典故以及诊疗过程的评论与总结,使查看医案者对此案例有更详细的了解。

每组核心元数据包括若干核心元素(本文表示为一级元素),“医案基本信息”和“患者基本信息”的核心元素参照 DC 元数据[24]的 6 个元素(Title、Creator、Date、Source、Publisher、Language)、CISMeF[25]元数据的 2 个元素(city、province or state)、《健康档案公用数据元标准》和《电子病历基本数据集标准》(WS445 – 2014)。“诊断过程”“诊断评定”和“诊疗措施”的核心元素均来自 WS 445.1 – 2014 电子病历基本数据集 – 第 1 部分、病历概要和 WS445.2 – 2014 电子病历基本数据集 – 第 2 部分、门(急)诊病历。“分析总结”的核心元素来自“传统中医师承病案模板”。为保证医案元数据元素的全面性与详细性,部分核心元素也可扩展,在每个需要详细化的一级元素下纵向扩展二级元素描述一级元素,若再有需要,可纵向扩展三级元素描述二级元素(表 2 中为二级元素和三级元素来源)。

表 2　中医医案元数据核心元素及扩展元素

核心元数据	元素来源	一级元素集	二级元素集	三级元素集
医案基本信息	《中医病历书写基本规范》	医案编号	无	无
	《电子病历基本架构与数据标准(试行)》	医案名称		
	《电子病历基本架构与数据标准(试行)》	医案来源		
		医案语言		
		医家		
		科别		
		就诊次数		
		就诊日期		
		发病节气		
患者基本信息	《中医病历书写基本规范》	患者姓名	无	无
	《电子病历基本架构与数据标准(试行)》	患者性别		
	《中医、中西医结合病历书写指南》	年龄		
		民族		
		职业		
		婚姻状况		
		籍贯		
		地址		
诊断过程	《中医、中西医结合病历书写指南》	体格检查	体温	无
			血压	
			辅助检查	
		主诉	无	无
	根据“《中医诊断学》– 问诊 – 问现病史”改编	现病史	主要症状	无
			患病时间	
			患病原因	
	文献:中医医疗与临床科研信息共享系统 –《中医临床术语字典》的构建		刻下症	
	《诊断学大辞典》		伴随症状	
	《中医诊断学》– 望诊	现病史	望诊	舌诊
				望神态
				望面色
				望形体
				望姿态
				望头面五官
				望皮肤毛发
				望躯体器官
				望四肢指趾
				望排出物
				望小儿指纹

（续表 2）

核心元数据	元素来源	一级元素集	二级元素集	三级元素集
	《中医诊断学》－闻诊		闻诊	闻气息
				闻气味
	《中医诊断学》－问诊		问诊	问寒热
				问汗
				问疼痛
				问头身
				问耳目
				问鼻喉
				问睡眠
				问饮食
				问口感与饮水
				问二便
				问经带
	《中医诊断学》－切诊		切诊	脉诊
				按诊
	《中医诊断学》－问诊		过往诊察治疗过程	无
	《中医诊断学》－问诊	既往史	平时健康情况	无
	《中医病历书写基本规范》		既往疾病史	
	《电子病历基本架构与数据标准(试行)》		手术史	
		既往史	过敏史	
			预防接种史	
	《电子病历基本数据集标准》		输血史	
	《中医诊断学》－问诊－生活史	个人史	生活习惯	无
	根据《中医诊断学》－问诊－生活史 添加的		个人经历	物理环境变化
				人文环境变化
	《电子病历基本架构与数据标准(试行)》		职业病史	无
	《中医病历书写基本规范》	家族史	家族遗传性疾病史	无
诊断评定	《中医、中西医结合病历书写指南》	初步诊断	中医诊断	无
			西医诊断	
		辩证分析	无	
诊疗措施	《中医、中西医结合病历书写指南》	治则	无	无
		处方		
		医嘱	每月剂量	
			采用剂型	
			用药方法	
			服用要求	
			饮食忌口	
分析体会	传统中医师承病案模板	按语	无	无

3.2 中医医案实例内容分析与增补元数据

通过文献分析，确定中医医案核心元数据与扩展元素后，为了防止有元数据缺失，运用内容分析法对网络中医医案实例内容进行解析，编码出每个具体的医案内容包含的元数据元素，与上文表 2 所得的元素进行对比，找到上文元数据没有覆盖到的有价值的诊疗信息元素。首先，通过网络调查，采集中医医案实例数据，抽取部分作为分析对象；再根据表 2 中元素设计分析类目及编码表；最后根据编码表解析网络中医医案实例，确定实例中无法归类的诊疗信息元素，进行元数据元素增补。

3.2.1 中医医案实例调研

本文的样本主要选自互联网中时常更新的中医医案相关网站，包括中医中药网、中医世家－医案心得、爱爱医网站、好大夫在线网站。这 4 个网站包含许多中医疾病与治疗处方，本文需要选取的样本是含有中医医案完整实例的文章，排除只讲述方法或只有部分案例的文章，最终平均每个网站选取符合需求的文章 50 篇，发布时间为 2019 年 10 月 20 日之前，从第一页顺序向后选取，共选取 200 篇。

3.2.2 设计类目与编码表

类目是内容分析法中对内容进行分类的项目，在

李璐，马捷，孙恒宇，等. 面向中医诊疗知识库的医案元数据模型构建研究[J]. 图书情报工作,2021,65(2):4-16.

本文中与核心元数据对应。共分为 6 个类目(见图3):医案基本信息、患者基本信息、诊断过程、诊断评定、诊疗措施、分析总结。

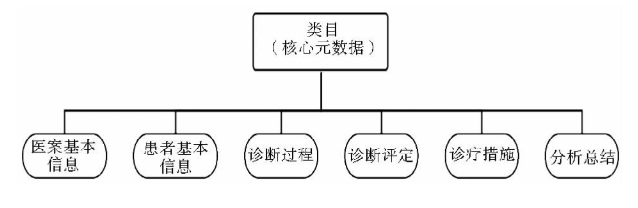


图3　类目(核心元数据)

参照上文所得元素对医案实例内容进行编码,每个编码表示为某一核心元数据的最后一级元素。例如,“诊断过程”的某项扩展:“现病史”——“切诊”——“脉诊”,“脉诊”为“诊断过程”的最后一级元素,编码表示为“脉诊”。本文中编码设计主要分为两个部分,一是实施预编码以检查编码员间的信度,二是正式实施内容编码。

(1)实施预编码以检查编码员间的信度。这里的信度可以定义为不同编码员对内容归类的一致性。为确保能够准确编码,需要对编码员之间的信度进行实际评估,本文将选择两位编码员,一位是笔者,另一位是研究组中与笔者研究方向类似的成员。预编码之前,笔者对另一位编码员进行了培训,说明各元素的涵义及可能对应的句子。接下来开始检验二者的信度,编码信度检验标准一般是在 95% 的置信区间内,编码员之间的总体一致性水平为 85% 以上,即表示编码操作通过信度检验[26]。

首先确定信度检验的样本个数。根据《内容分析法:媒介信息量化研究》[27]一书中提出的不同总体规模检验信度所需要的样本量的理论,本文总共选取 200 篇医案实例样本,在 95% 的置信区间内,假设编码员总体的一致性水平为 90%,根据公式(1),需要随机抽取 68 篇样本对编码员信度进行检验。

$$n = \frac{(N-1)(0.03)^2 + PQN}{(N-1)(0.03)^2 + PQ} \qquad \text{公式(1)}$$

其中,n = 信度检验的样本量,N = 研究中的样本量,P = 总体的一致性水平,Q = 1 - P。

接下来,让两位编码员对这 68 篇医案实例内容进行编码,如双方有不确定的编码情况,将句子记录下来,由第三人(中医专业的学生)参与分析后确定如何编码赋值。编码后的结果为:N1 = 93,N2 = 91,M = 85,n = 2。代入信度检验公式,如公式(2)所示:

$$R = \frac{n \times K}{1 + (n-1) \times K} \qquad K = \frac{2M}{N_1 + N_2} \qquad \text{公式(2)}$$

其中,n 为编码员数量,M 为两者都完全同意的元数据元素数,N_1 为第一编码员分析的编码数,N_2 为第二编码员分析的编码数。K 为两个编码员之间的相互同意度。

根据以上公式得到两位编码员之间的相互同意度为 K = 92%,信度 R = 96%,其结果大于 90%,达到了假设的总体一致水平 90% 以上,通过信度检验,可以继续进行编码。

(2)正式实施内容编码。通过信度检验后,对剩下的医案实例继续编码,同预编码一样,双方如有不确定的编码情况,将句子记录下来,由第三人(中医专业的学生)参与分析后确定如何编码赋值,最终得到表 3 所示的编码表。

3.3　信度检验

信度是指测验结果的准确与可靠的程度[28],一般会通过对测量值与评判值之间的相差程度,判断是否可以采用测量后得出的结论。为了检验内容分析所得元素的可靠性,本研究又选择两位研究组其他成员作为评判者进行元数据元素的分析,首先将整理好的编码表交给两位评判者,对两位同样进行编码前的培训,说明编码方法及各元素的涵义与示例,并让两位尝试编码 5 篇相同医案,检验其是否掌握编码方法。在笔者认定两位评判者已掌握编码要领后,在上文的 4 个网站中又抽取 40 篇医案让两位评判者共同进行信度检验,将得出的结果代入公式(2),得到两位评判者之间的相互同意度为 K = 0.85,信度 R = 0.92,为可以接受的信度标准,表明编码表可以使用。

3.4　确定增补元数据元素

通过对医案实例的内容分析,得到编码表(见表3),与表 2 中元素对比有新补充的元素(两位编码员共同商议应该归为某一核心元数据的某一级元素,结果如表 4 所示,表中灰尘色阴影字体为新增补的元素)。有一项元素未在网络调研的案例中发现使用:诊断过程——现病史——望诊——望小儿指纹。由于还会通过实地调查进一步优化元素,暂时保留这一元素,在实地调查之后再决定是否删除。

4　中医医案元数据元素优化与模型建立

4.1　元素的验证与优化过程

依据文献分析和和网络医案实例内容分析,还不能完全确认得到的元素是否能与实际诊疗过程相匹配,需要进一步通过实地调查法对医案元数据元素进行验证与优化。实地调查法是研究者获取研究资料的

表 3 编码表

医案基本信息						
医案编号	医案名称	医案来源	医案语言	医家	科别	就诊次数
就诊日期	发病节气					
患者基本信息						
患者姓名	患者性别	年龄	民族	职业	婚姻状况	籍贯
地址						
诊断过程						
体温	血压	辅助检查	主诉	主要症状	患病时间	患病原因
刻下症	用药后情况	伴随症状	舌诊	望神态	望面色	望形体
望姿态	望头面五官	望皮肤毛发	望躯体器官	望四肢指趾	望排出物	闻气息
闻气味	问寒热	问汗	问疼痛	问周身	问耳目	问鼻喉
问睡眠	问饮食	问口感与饮水	问二便	问经带	问吸烟情况	问饮酒情况
问运动情况	问婚育情况	问服药情况	问呼吸消化	脉诊	按诊	过往诊察 治疗过程
平时健康情况	既往疾病史	既往相同病史	手术史	过敏史	用药史	预防接种史
冶游史	输血史	生活习惯	物理环境变化	人文环境变化	心态情绪	精神
工作、生活压力	职业病史	家族遗传性疾病史				
诊断评定						
中医诊断	西医诊断	辩证分析				
诊疗措施						
治则	处方	医嘱－每日剂量	医嘱－采用剂量	医嘱－用药方法	医嘱－服用要求	医嘱－饮食忌口
医嘱－心态	医嘱－运动	医嘱－其他				
分析总结						
按语						

表 4 增补元数据元素集

核心元数据	一级元素集	二级元素集	三级元素集
诊断过程	现病史	用药后情况	
		问诊	问吸烟情况
			问饮酒情况
			问运动情况
			问婚育情况
			问服药情况
			问呼吸消化
	既往史	既往相同病史	
		用药史	
		冶游史	
	个人史	精神状态	心态情绪
			工作、生活压力
			精神
诊疗措施	医嘱	心态	
		运动	
		其他	

最基本的途径,它强调在实际社会生活中的亲自观察和访问,最大优势在于它的直观性和可靠性。实地调查法的形式多种多样,主要包括现场直接观察法、访问法、实验法等方法[29]。然而中医的诊病过程部分是只可意会不可言传的,经验充足的医生在诊病过程中的语言往往比较简洁,不会将诊病的流程尽数表达,研究者在观察过程中出现疑问之处会通过访问的方式请求医生解答,以免错过某个元素的选择。为此,本文采取实地调查法中的现场直接观察法和访问法,对初步优化的元素进一步优化,以求保证元素的全面性与可靠性。

实地调查的过程可分为准备、现场调查和整理资料三个阶段:

(1)准备阶段。在去中医诊疗现场之前,将上文中通过标准规范和医案实例内容得到的元素整理成表格的形式,打印成元素表文本,供实地调查记录。记录人员为上文的两位编码员,即笔者和研究组中与笔者研究类似的成员。

(2)现场调查阶段。本研究选择长春市与松原市的 10 位中医为实地调查对象,在调查开始之前,对医生讲述此次调查的目的是了解中医在整个诊病过程中具体做了什么,研究者需在一旁观察并记录诊病过程中医生的行为和与患者交谈关于病情的详细内容,记录内容均与病情相关,不会涉及病人隐私。同时将准备阶段打印的元素表文本交给医生查看,以期医生可以对所得到的元素的合理性和实用性给出相应的意见。

在观察过程中,两位编码员根据元素表中的元素记录诊病过程,将出现疑问的地方标记出来等待休息时间商议。在休息时间,研究者会整理出现疑问的地方,先内部商议可以解决的部分,统一二者在观察过程中产生的需要医生回答的疑问之处,再对当下接受实地调查的医生进行访问,以期对该医生诊疗过程的方式有所了解,对所记录的文字能够有更充分的理解。访问问题(即研究者在观察过程中存在疑问的问题,部分问题会问几个医生,回答中为保护医生隐私用医生A、B、C表达)与回答如表5所示:

表5 实地调查访问问题及回答

序号	访问问题	医生回答
1	为什么您经常询问患者的身高与体重,我之前看医案的时候并未见其有所涉及?	患者的身高与体重也是开药与用药剂量的评判因素
2	在诊断过程中的时候,会看患者的面色之类的吗?	医生A:会的,患者一进来或者一坐在凳子上的时候就会观察他们的面部,扫一眼就看的差不多
		医生B:肯定会,这个是所有中医的一种习惯,进门必看
		医生C:会看,这是中医望诊里面的一项,虽然我不说出来,但是中医的“望闻问切”我们在诊病过程中缺一不可
3	之前有一位老人带着帽子看诊,您让他摘掉帽子,是要看头发吗?	是的,一般病人在看中医的时候,最好不要带口罩帽子等遮盖物,女子最好不要化妆,这样方便观察头部以及面部真实的模样
4	之前听您说过病人大喘气,但是您说的比较少,之前我在其他中医处时也较少听到,平时在诊病过程中会注意患者的气息之类的吗?	会的,这个在中医中属于“闻诊”中的一项,“闻诊”主要包括听声音和闻气味,看病看的久了,这些都是习惯性做的事情,注意病人的呼吸、语气之类的,身上散发出来的气味等,看诊过程中可能不会说出来,但是“闻诊”作为中医四大诊察方法之一,都是必须做的
5	看您有时会握病人的手并查看,是感受手冷热吗?	一是为了看手冷热,手足冷或发热都是诊断病症的一项因素;二是通过对手部的观察,也可以观察到身体内部的状态
6	为什么您经常询问病人晚上睡觉是否做梦,做梦是否可以记住,有时还会讲述梦境等?	有些人经常做梦可能习以为常,但是从同一个人的做梦情况可以探究他的睡眠质量,精神状态等,有时还可以探究他的生活境况等。例如有些人会做关于考试的梦,可能就会猜测最近学习、工作或生活压力大导致,就可以询问病人是否有什么压力之类的
7	为什么您有时会敲患者的肚子,这个属于按诊吗?	不属于,如果是按诊的话,我不会选择敲击的方式,而是直接按压肚子或其他部位;敲肚子,是想听敲击的声音,去判断是否有病症存在或病症的发展程度
8	您这里是儿科,几乎每个小患者来的时候都会使用听诊器,但是现在是流感爆发季,是否平时会少用一些?	不会的,现在越来越多的家长会带着孩子来看中医,虽然有一些是特殊病症,或者调理脾胃的,但是大多数孩子都是因为发烧,感冒等来看医生,有许多孩子都是在医院打了吊针以后再来看中医的,而孩子发烧感冒最害怕的就是肺部出现问题,并且听诊器会听孩子的心跳,也是对孩子病情产生影响的一项。所以大多数来这里看病的孩子都会用上听诊器,先确定或排除肺部和心脏的问题

(3)整理资料阶段。在调研第一位医生时,医生在查看过元素表后指出“医案基本信息”和“患者基本信息”中的元素都是病历系统自动生成的,并让笔者查看了病历系统确定。

两位编码员在对每个医生实地调查之后都会一起对记录的内容进行整理,整理出现元素的个数,并且统一有异议的部分,商议需要补充的元素。共得到972条医案案例,整理出如表6所示的元素(同编码表的表示方式相同,为某一核心元数据的最后一级元素;表格内有灰色阴影部分为新增补元素)。

表6中的“约束”一项,为共享元数据标准注解中的“约束/条件”,是说明元数据元素是否必须选取的属性,M为必选,表明该元数据元素必须选择;O为可选,是根据实际应用可以选择也可以不选的元数据元素;C为条件必选,是当条件满足时,元数据元素为必选。

核心元素皆为必选元素,“医案基本信息”和“患者基本信息”的17个核心元素以及“主诉”“治则”“处方”“辩证分析”皆为核心元素,除此之外必选元素还有11个,其中“主要症状”“患病时间”“问二便”“中医诊断”“医嘱-每日剂量”“医嘱-采用剂型”“医嘱-用药方法”7个元素皆是100%存在的,而“舌诊”“问睡眠”“问饮食”“脉诊”4个元素虽未达到100%,但都在95%以上,可见其在诊疗过程中的必要性,以上皆是中医医案元数据必须选择的元素。

可选元素共59个,其中“用药后情况”“问寒热”“问汗”“问周身”“问口感与饮水”“问呼吸消化”“医嘱-服用要求”“医嘱-饮食忌口”“医嘱-心态”“医

图书情报工作
第 65 卷 第 2 期　2021 年 1 月

表 6　实地调查优化的元素

元素	词频	比例/%	约束	元素	词频	比例/%	约束	元素	词频	比例/%	约束
医案编号	972	100	M	望姿态	26	2.67	O	平时健康情况	328	33.74	O
医案名称	972	100	M	望头面五官	68	7	O	既往疾病史	492	50.62	O
医案来源	972	100	M	望皮肤毛发	65	6.69	O	既往相同病史	226	23.25	C
医案语言	972	100	M	望躯体器官	178	18.31	O	手术史	124	12.76	O
医家	972	100	M	望四肢指趾	478	49.18	O	过敏史	208	21.4	O
科别	972	100	M	望排出物	86	8.85	O	用药史	320	32.92	O
就诊次数	972	100	M	望小儿指纹	0	0		预防接种史	94	9.67	O
就诊日期	972	100	M	闻气味	12	1.23	O	冶游史	2	0.21	O
发病节气	972	100	M	闻气息	68	7	O	输血史	16	1.65	O
患者姓名	972	100	M	叩鸣	378	38.89	O	外伤史	26	2.67	O
患者性别	972	100	M	问寒热	862	88.68	O	生活习惯	18	1.85	O
年龄	972	100	M	问汗	824	84.77	O	物理环境变化	102	10.49	O
民族	972	100	M	问疼痛	488	50.21	O	人文环境变化	82	8.44	O
职业	972	100	M	问周身	682	70.16	O	心态情绪	562	57.82	O
婚姻状况	972	100	M	问耳目	366	37.65	O	工作、生活压力	436	44.86	O
籍贯	972	100	M	问鼻喉	348	35.8	O	精神	168	17.28	O
地址	972	100	M	问睡眠	970	99.79	M	职业病史	2	0.21	C
体温	112	11.52	O	问饮食	964	99.18	M	家族遗传性疾病史	84	8.64	C
血压	235	24.18	O	问口感与饮水	820	84.36	O	中医诊断	972	100	M
听诊器	98	10.08	O	问二便	972	100	M	西医诊断	0	0	C
辅助检查	675	69.44	C	问经带	578	59.47	O	辩证分析	972	100	M
主诉	972	100	M	问吸烟情况	156	16.05	O	治则	972	100	M
主要症状	972	100	M	问饮酒情况	178	18.31	O	处方	972	100	M
患病时间	972	100	M	问运动情况	366	37.65	O	医嘱－每日剂量	972	100	M
患病原因	342	35.19	O	问婚育情况	498	51.23	O	医嘱－采用剂量	972	100	M
刻下症	357	36.73	O	问服药情况	378	38.89	O	医嘱－用药方法	972	100	M
用药后情况	602	61.93	O	问呼吸消化	673	69.24	O	医嘱－服用要求	782	80.45	O
伴随症状	524	53.91	O	问身高体重	482	49.59	O	医嘱－饮食忌口	896	92.18	O
舌诊	958	98.56	M	问精神	358	36.83	O	医嘱－心态	679	69.86	O
望神态	56	5.76	O	脉诊	942	96.91	M	医嘱－运动	682	70.16	O
望面色	36	3.7	O	按诊	268	27.57	O	医嘱－其他	489	50.31	O
望形体	138	14.2	O	过往诊察							
治疗情况	352	36.21	O	按语	0	0	M				

嘱－运动”10 个元素皆在 60% 以上，是中医诊疗过程中较为重要的元素。

属于条件必选（C）的元素共有 5 个：“辅助检查”的条件是含有各类医疗项目的检查结果；“既往相同病史”的条件是以前有相同的疾病病情或相似病历的情况；“职业病史”的条件是正在从事或以前从事会影响身体或疾病的工作，特别是放射性工作；“家族遗传性疾病史”的条件是家族中人有遗传性疾病的情况。

“西医诊断”虽然在实地调查中并无展现,但是在医案实例中有部分医案会有此项元素,因此归为条件必选元素,条件是有西医参与诊断,或存在西医诊断的病名。

4.2 元素优化结果

中医理论来源于实践,实地调查得出的结论是对中医文献内容的补充。根据对实地调查的记录整理得出的元素(见表6),与前文医案实例内容分析所得元素(见表3)相比有所增减,增添的元素(见表7)及增添原因如下:

表7 实地调研增补元数据元素集

核心元数据	一级元素集	二级元素集	三级元素集
诊断过程	体格检查	听诊器	
	现病史	闻诊	叩鸣
		问诊	问身高体重
			问精神
	既往史	外伤史	

(1)听诊器。“听诊器”是医生常用的诊听患者内部器官的工具,由于本次实地调查中的一位中医为儿科中医,经常使用听诊器听儿童的心肺部,几乎所看儿科病例有90%以上含“听诊器”一项,因此将其纵向扩展到一级元素“体格检查”的二级元素中。

(2)叩鸣。《中医诊断学》中表明“闻诊”包括“听声音”和“嗅气味”两个方面内容。医案中一般表现出来的“听声音”,是指听患者的气息,包括喘气声、咳嗽声、叹息声等。但是实地调查发现,部分中医会通过敲击身体某个部位(例如:腹部)发出的声音去判断患者的病情,笔者与评判者商议将这种情形定义为“叩鸣”,扩展到二级元素“闻诊”的三级元素中。

(3)问身高体重。身高体重一般与病情无关,医案中并无表述。但是中医医生会根据患者的身高体重考虑患者的用药量,需要在二级元素“问诊”中纵向扩展“问身高体重”三级元素。

(4)问精神。是指询问患者患病之后的精神状态。在实地调查过程中,有患者表示患病后心情烦躁、睡眠质量不高等,但平时生活压力不大。可以注意到精神状态有患病后与平日之间的区别,需要将二者分开表示,在二级元素“问诊”中纵向扩展“问精神”三级元素,应注意与“个人史”中的“精神状态”有所区别。

(5)外伤史。是指患者曾经有过明确可查的受伤的情况。在实地调查过程中,有患者叙述以前曾经出过严重的车祸,或曾经骨折过等,笔者参照《入院记录书写要求及格式》,拆分“既往史”中的“手术外伤史”,对本文的一级元素“既往史”纵向扩展二级元素“外伤史”。

经过医案实例的调研和实地调查两次优化元素的过程,“诊断过程——现病史——望诊——望小儿指纹”这一项都没有在调查结果中出现,因此经过笔者与研究组成员商议,决定删去“望小儿指纹”,如果在未来录入数据时出现,再予以添加。

4.3 元数据模型建立

经过上述过程对中医医案元数据元素的优化,最终确立了中医医案元数据模型(见图4)。该元数据模型涵盖完整的中医诊疗过程,明确医案内容的格式,减小了医案信息共享与获取的屏障,便于医案信息的保存、管理与检索。

本文的中医医案元数据模型对中医诊疗过程中的关键内容揭示得更为全面和深入,对比已有中医医案标准和规范研究,可以更好地表达诊疗关键细节。中医诊断过程包含了大量的中医个人经验(隐性知识),平时病历中展示很少,可能会导致中医诊疗过程的知识挖掘并不全面。本文通过参照以往标准规范与元数据模型已有的信息,通过实际病历分析和医疗实地调研等规范的研究方法,科学论证和构建出中医医案元数据模型,以期加大医生隐性知识显性化的几率,使医案资源能够得到更好的开发和利用。

5 结语

中医是我国珍贵的文化遗产,中医医案作为中医文化传承的纽带,对中医的继承与发展至关重要。然而目前互联网中的医案信息共享受阻,检索不精确,为医案内容提供统一的格式成为首要考虑的重点。本文采用理论与实践相结合的思想设计中医医案元数据,首先参照中医标准规范与元数据标准等方面的文献,得到核心元素并进行元素扩展;其次搜集中医医案实例信息,研究中医医案实例内容数据的特点,通过内容分析法对医案内容进行分析,初步界定医案元数据元素;再通过实地调查法,对中医诊疗过程进行观察和访问,优化医案元数据元素;最后建立元数据模型。本文构建的元数据模型精度较高,可以将诊疗过程中的信息进行深度结构化组织,有利于诊疗数据库或者知识库的多维度检索和知识挖掘,深化中医诊疗知识组织理论研究与实践工作。

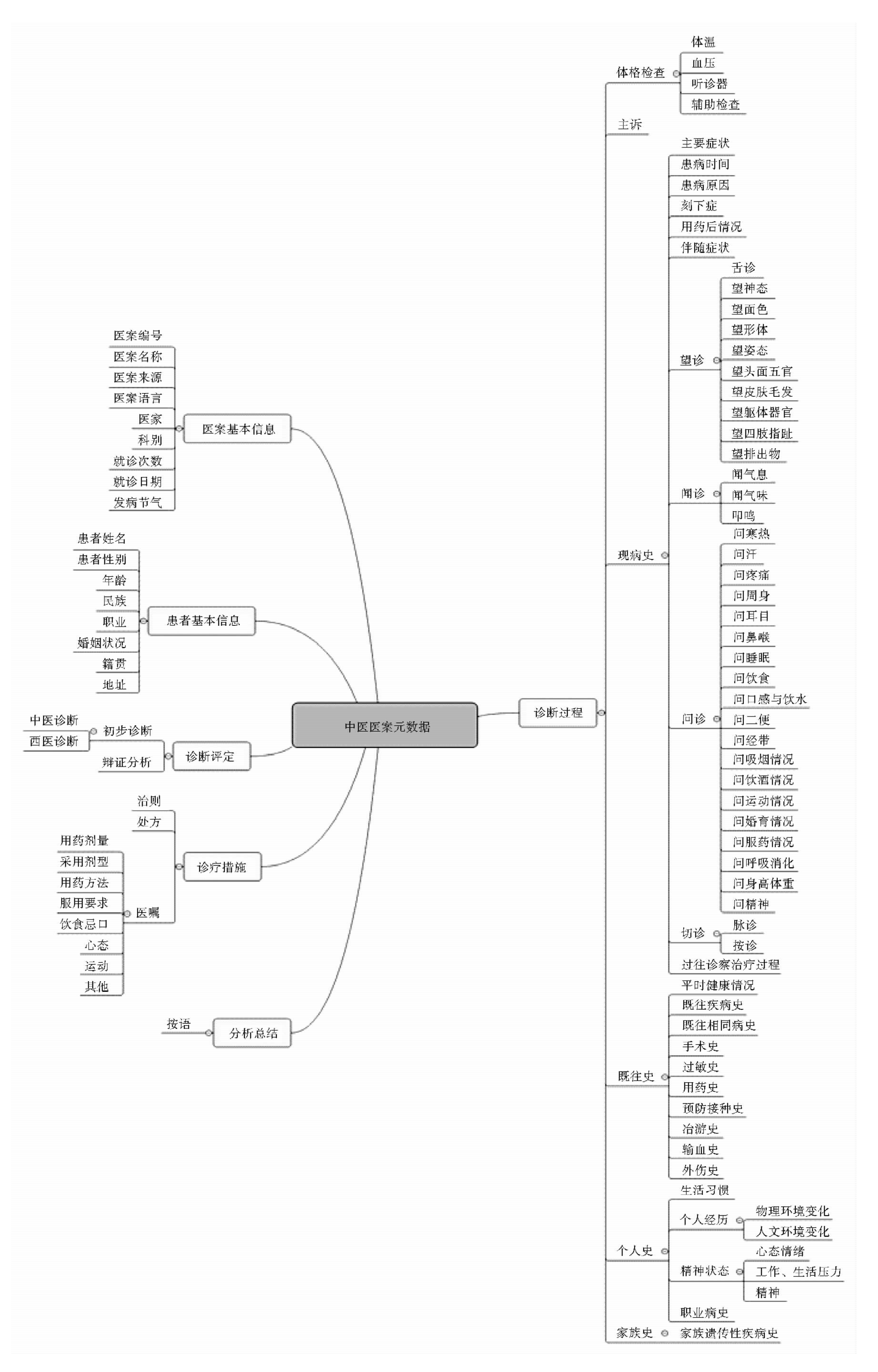


图 4　中医医案元数据模型

参考文献：

[1] 崔俊波，陈宝贵. 名中医传承工作室传承思路及经验探讨[J]. 医学信息，2015(10):273－273.

[2] 任廷革，刘晓峰，高剑波，等. “中医药基础数据库系统”介绍[J]. 中国中医药信息杂志，20018(11):90－91.

[3] 吴中平，何新慧，柯雪帆，等. 医案查询统计分析系统介绍[J]. 上海中医药杂志,2003,37(3):54－56.

[4] 于琦，崔蒙，李园白，等. 中医药文献数据库建设规范研究[J]. 世界科学技术－中医药现代化,2014,16(11):2304－2307.

[5] 苏礼. 中医医案学概论[M]. 北京：人民卫生出版社，2009.

[6] 中医研究院. 中医名词术语选释[M]. 北京：人民卫生出版社，1973.

[7] 高殿璞，王映辉，张润顺，等. 中医医案规范化研究述评[J]. 中国中医药信息杂志,2018,25(5):131－135.

[8] 孙光荣. 医案研究与撰写的思路与方法[J]. 北京中医药大学学报(中医临床版),2013,20(5):3－6.

[9] 杨建宇，李彦知，范竹雯，等. 浅谈医案的写作与编辑——兼论中医大师孙光荣教授研究中医典型医案的学术思想[J]. 光明中医,2011,26(11):2369－2373.

[10] 李纪宾，邹小农，乔友林. 元数据及其在医学科学数据共享中的应用[J]. 医学情报工作,2006,27(1):10－14.

[11] 李海燕，崔蒙，任冠华，等. ISO/TC215 传统医学信息标准化工作进展[J]. 国际中医中药杂志,2011,33(3):193－195.

[12]《中国中医药报》. ISO 首发两项中医药信息国际标准——《中医药学语言系统语义网络框架》和《中医药文献元数据》[J]. 江苏中医药，2014,46(10):27－27.

[13] Clinical Trial Electronic Data Capture Task Group. Electronic clinical data capture position paper revision 1[EB/OL]. [2020－05－14]. https://rammellconsulting. files. wordpress. com/2011/05/edc-position-paper-revision-1-final-version. pdf.

[14] 李庚，李晓彦，温泽淮. 临床数据交换标准协会标准在中医药临床研究病例报告表设计中的应用[J]. 广州中医药大学学报，2014,31(1):138－141.

[15] 武烨，马路. 中医医案数据库的研发现状及展望[J]. 中华医学图书情报杂志,2012,21(11):49－52.

[16] 宋扬文，田瑞，陈进，等. 中医医案数据库的利用及需求调查研究[J]. 医学信息学杂志,2013,34(7):63－66,89.

[17] 田瑞，马路. 中医病案数据库元数据方案的设计[J]. 中华医学图书情报杂志,2014,23(10):66－69.

[18] 刘宝杰，陈进，马路. 中医病案数据库元数据结构设计实践[J]. 医学信息学杂志,2013,34(12):70－73.

[19] 翁子扬，李卓然. 艺术品数字化元数据模型构建[J]. 数字图书馆论坛,2019(6):11－18.

[20] 贾君枝，史璇. 数字博物馆元数据标准构建研究[J]. 山西大学学报(哲学社会科学版),2015,38(1):114－119.

[21] 李利. 地质灾害应急信息资源元数据模型构建研究[D]. 南昌：南昌大学,2014.

[22] 诸云强，项文新，冯敏，等. 可扩展的环境保护档案元数据研究与实践[J]. 档案学通讯，2012(5):78－82.

[23] 易军凯，周育彬，陈刚. 可扩展的数字博物馆元数据规范研究与实践[J]. 数字图书馆论坛，2014(2):43－53.

[24] Dcmi Usage Board. Dublin core metadata element set, version 1.1: reference description[EB/OL]. [2020－05－14]. https://www. dublincore. org/specifications/dublin-core/dces/

[25] CISMeF Team. CISMeF: catalog and index of French-language health internet resources. a quality-controlled subject gateway[EB/OL]. [2020－05－10]. http://www. chu-rouen. fr/cismef/cismefeng. html.

[26] 丁小建. 航空企业微博内容分析及其对用户关注行为影响研究[D]. 上海：东华大学,2014.

[27] 里夫，赖斯，菲克，等. 内容分析法：媒介信息量化研究技巧[M]. 嵇美云，译.2 版. 北京：清华大学出版社，2010.

[28] 时蓉华. 社会心理学词典[M]. 成都：四川人民出版社,1988.

[29] 胡镜清. 实地调查应该成为中医基础理论研究的基本方法之一[J]. 中国中医基础医学杂志，2014,20(12):6－8.

作者贡献说明：

李璐：数据调研及整理，论文撰写及修改；

马捷：论文选题制定，研究框架设计，论文修改及定稿；

孙恒宇：讲解中医知识，给予分析结果意见，修改论文；

王珏：数据收集与分析。

The Construction of Medical Case Metadata Model for the Knowledge Base of Traditional Chinese Medical Diagnosis and Treatment

Li Lu[1] Ma Jie[1,2] Sun Hengyu[3] Wang Jue[1]
[1] School of Management, Jilin University, Changchun 130022
[2] Information Resourses Research Center, Jilin University, Changchun 130022
[3] The First Clinical Hospital of Jilin Academy of Traditional Chinese Medicine, Changchun 130022

Abstract: [**Purpose/significance**] In the process of traditional Chinese medical record management, there are a lot of problems such as inconformity and insufficient accuracy of the traditional Chinese medical record metadata, which is not conducive to the information retrieval, knowledge sharing and mining of the traditional Chinese medical record knowledge base. This research establishes a standardized traditional Chinese medical record metadata model, which plays an important role in the deep multi-dimensional retrieval, knowledge sharing and mining of traditional Chinese medical records. [**Method/process**] This paper summarized the relevant standards and metadata standards of current Chinese medicine through literature survey method; by collecting examples of traditional Chinese medicine cases through network survey method, analyzed the contents of medical cases in conjunction with content analysis method, and initially defined the metadata of medical cases elements; then through the field investigation method, observed and visited the Chinese medicine diagnosis and treatment process at the Chinese medicine diagnosis and treatment site, optimized the preliminary defined medical record metadata elements, and obtained the final medical record metadata element set. [**Result/conclusion**] A medical record metadata model oriented to the knowledge base of Chinese medicine diagnosis and treatment is constructed to fully describe the diagnosis and treatment process of Chinese medicine and provide a reference for the unified metadata of Chinese medicine treatment records.

Keywords: metadata metadata model traditional Chinese medical record traditional Chinese diagnosis and treatment knowledge base traditional Chinese medical knowledge organization